

Variable-Scale Smoothing and Edge Detection Guided by Stereoscopy

Clark F. Olson

Jet Propulsion Laboratory, California Institute of Technology
Mail Stop 107-102, 4800 Oak Grove Drive, Pasadena, CA 91109
<http://robotics.jpl.nasa.gov/people/olson/homepage.html>

Abstract

It is typical in computer vision applications to examine a single scale or to consider some space of scales in the image without knowing which scale is appropriate for each location in the image. However, many images contain a wide variation in the distance to the scene points, and thus objects of the same size can appear at greatly differing scales in the image. We present a method where the scale of the smoothing and edge detection is varied locally according to the distance to the scene point, which we estimate through stereoscopy. The edges that are detected are thus at the same scale in the world, rather than at the same scale in the image. This method is implemented efficiently by smoothing the image at a discrete set of scales and performing interpolation to estimate the response at the correct scale for each pixel. The application of this technique to an ordnance recognition problem has resulted in a considerable improvement in performance.

1 Introduction

Image smoothing and edge detection have been intensely studied subjects in computer vision and image processing. The selection of an appropriate scale for these processes is a problem that has received less attention. It is well known that using a single fixed scale over the entire image often produces undesirable results, since edge phenomena occur at a multitude of scales. To alleviate this problem, techniques that examine the entire space of scales [6, 7, 15] or that adaptively select a scale based on local image properties [5, 8, 11] have been developed. However, the optimal method for combining the information from the scale-space is unclear, and the scale selection methods base their decisions on image properties, rather than the true scale at which the phenomena occur.

In many applications, it is desirable to detect edges that are at the same scale in the world, which we call

the *true scale*, rather than at the same scale in the image or by selecting a scale based on local image properties. Consider, for example, an image containing a textured surface in the foreground and a structure of interest further from the camera. Techniques based on local image properties consider the textured surface at the scale it appears in the image. At this scale, the edges may appear quite significant, while, in fact, this appearance is only due to perspective effects. If a method (such as stereoscopy) is available to determine the distance of the scene points from the camera, we can safely smooth these phenomena, while preserving the significant edges. Furthermore, if we seek objects of known size, the smoothing and edge detection process can be tuned to detect edges of the appropriate scale, regardless of their distance from the camera.

In addition to its value for scale selection, the stereoscopy output is useful in determining edge salience with respect to the scene characteristics. For example, edge salience measures such as length and straightness have been used [14]. However, the values these measures take are highly dependent on the distance of the edge from the camera. The stereo range information can be used to normalize these measures with respect to scene size and it is thus possible to determine edge salience with respect to the true scale rather than the image scale.

We have implemented these techniques as a variation of the Canny edge detector [1], but they can be applied to most edge detection methods. A mapping function between the local depth at each pixel and the image scale values is first determined. We next smooth the image at a discrete set of scales using Gaussian derivative filters. The appropriate scale response at each pixel is then interpolated from the discrete set of filter responses (similar to idea of steerable or deformable filters [2, 13]). These responses are next normalized, since the overall response to a Gaussian derivative filter is a function of the scale of the filter. Edge detection then proceeds normally, extracting edge chains using non-maxima suppression

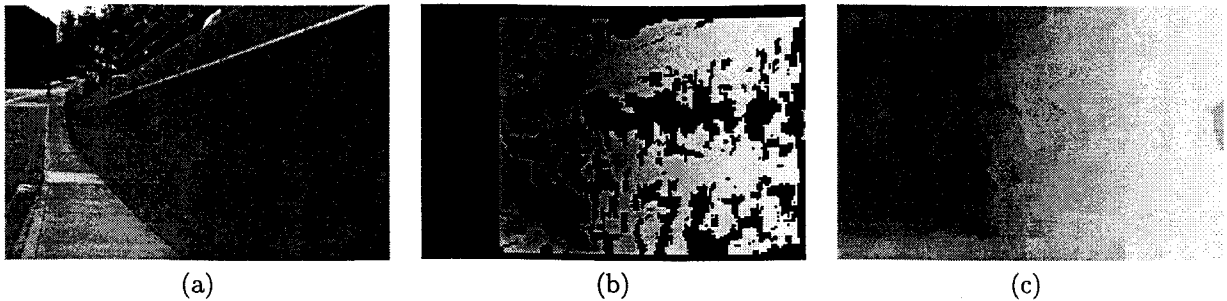


Figure 1: Range data extracted from a stereo pair. (a) Left image of a stereo pair. (b) Distance from the camera mapped into gray values. Black pixels indicate no valid range data. (c) Distances after filling pixels with no range data.

and hysteresis thresholding. These edge chains are finally passed to a stage that determines edge salience with the help of the stereo range map.

We give results that show that these techniques result in significantly improved performance in a target recognition application in which unexploded ordnance is detected using the image edge map.

2 Depth acquisition

While any method that can associate depth values with image pixels could be used with this method, we concentrate on the use of stereoscopy to compute dense range maps of the scene. The techniques that we use to compute the stereo range data have been described elsewhere [9, 10]. We briefly summarize this method here.

An off-line step, where the stereo camera rig is calibrated, must first be performed. We use a camera model that allows arbitrary affine transformations of the image plane [16] and that has been extended to include radial lens distortion [3]. The remainder of the method is performed on-line.

At run-time, each image is first warped to remove the lens distortion and the images are rectified so that the corresponding scan-lines yield corresponding epipolar lines in the image. The disparity between the left and right images is measured for each pixel by minimizing the sum-of-squared-difference (SSD) measure of windows around the pixel in the Laplacian of the image. Subpixel disparity estimates are computed using parabolic interpolation on the SSD values neighboring the minimum. Outliers are removed through consistency checking and smoothing is performed over a 3×3 window to reduce noise. Finally, the coordinates of each pixel are computed using triangulation.

Note that not every pixel is assigned a range with this method. There are a number of factors that result

in various pixels not being assigned a range, including occlusion, window effects, finite disparity limits, low texture, and outliers. Despite this problem, we must have a range estimate at each point in the image in order to estimate the scale that should be used for smoothing at that point. To resolve this dilemma, we propagate the range values from neighboring pixels using a simple method that approximates nearest neighbor search.

Figure 1 shows an example of the range data computed using these techniques. In this case, we fail to get range data at the left edge of the image, since this is the left image of a stereo pair, and there are significant areas over the rest of the image where the range data is discarded as not reliable. These values are filled with good estimates using the propagation techniques.

3 Smoothing with variable scale

We perform variable-scale smoothing using the stereo range data to select the appropriate scale at each pixel. The first step is to construct a mapping between the range data that has been computed for the scene and the scale at which the smoothing should be performed. We construct this mapping off-line prior to the smoothing. However, this mapping could be easily constructed on-line in order to allow it to vary with scene parameters.

We use the following mapping function:

$$\sigma(x, y) = \frac{K}{R(x, y)},$$

where $R(x, y)$ is the range computed at the image point (x, y) , $\sigma(x, y)$ is the scale to be used at (x, y) , and K is a pre-determined constant.

The constant, K , in this mapping function can be determined using several methods. One possibility is

to modify an automatic scale selection method (see, for example, [4]) to examine the image scale normalized by the depth values. A second possibility is to not limit ourselves to a single scale, but to consider the scale-space as in [15]. In this case, the scale-space can be warped such that the scales levels correspond to true scale rather than image scale. We use a third alternative. Since our primary application for these techniques is in detecting objects of known size, we select K based on the known size of the objects.

In following with Canny's edge detection method [1], we use Gaussian filters to perform image smoothing. However, since we vary the scale at each pixel, the responses we want are generated by:

$$S_x(x, y) = \sum_{i=-W}^W \sum_{j=-W}^W I(x+i, y+j) N_{\sigma(x,y)}(i, j),$$

where $I(x, y)$ is the image brightness at (x, y) , $N_{\Sigma}(X, Y) = \frac{1}{\Sigma\sqrt{2\pi}} e^{-\frac{X^2+Y^2}{2\Sigma^2}}$, and $2W+1$ is the filter window size. Unfortunately, it is not efficient to compute this exactly for each image pixel. We perform this unconventional operation by convolving the image with a discrete set of Gaussian filters of various scales and interpolating the result at the appropriate scale for each pixel. This method for approximating a continuum of parameterized filters is similar to the techniques of steerable filters [2] and deformable kernels [13]. However, we have chosen parabolic interpolation rather than the linear combinations of the deformable kernels technique for simplicity and ease of implementation.

Since the range of scales that we are concerned with may be very large and Koenderink [6] has shown that a logarithmic sampling of the scale space is stable and in accordance with the principle that no scale should be preferred above others, we work in the $\log_2 \sigma$ domain. We have found that using discrete scales related by factors of two ($\sigma_n = 2^n \sigma_0$) is both convenient and effective.

The result of smoothing at each pixel with a filter of scale $\sigma(x, y)$ can be estimated through parabolic interpolation using the response of the discrete filter that is closest to the desired scale, $F_{\sigma_k}(x, y)$, and its two neighbors, $F_{\sigma_{k-1}}(x, y)$ and $F_{\sigma_{k+1}}(x, y)$. In determining an equation that yields the appropriate response, it is useful to perform a coordinate transform such that $z = \log_2 \frac{\sigma(x,y)}{\sigma_k}$. For $\sigma_{k-1} = \frac{1}{2}\sigma_k = \frac{1}{4}\sigma_{k+1}$, this yields $z_{k-1} = -1$, $z_k = 0$, and $z_{k+1} = 1$. With this transformation it is simple to show that the response

we want is given by:

$$F(x, y) \approx az^2 + bz + c \quad (1)$$

$$a = \frac{1}{2}(F_{\sigma_{k-1}} - 2F_{\sigma_k} + F_{\sigma_{k+1}}) \quad (2)$$

$$b = \frac{1}{2}(F_{\sigma_{k+1}} - F_{\sigma_{k-1}}) \quad (3)$$

$$c = F_{\sigma_k} \quad (4)$$

$$z = \log_2 \frac{\sigma(x, y)}{\sigma_k} \quad (5)$$

4 Edge detection

Following the variable-scale smoothing described above, we proceed with Canny's edge detection method [1] on the smoothed image. This technique computes the image gradients over the image in the x - and y -directions in order to determine the orientation and magnitude of the gradient at each pixel. Note, however, that if the gradient magnitudes are to be comparable, we must normalize them. This can be easily seen by noticing that the response of a step edge to a Gaussian derivative filter varies with the scale of the filter. A 1-dimensional Gaussian derivative aligned with a step edge yields a response proportional to $\frac{1}{\sigma}$. To correct this problem, we normalize the gradient magnitude at each pixel by multiplying by $\sigma(x, y)$.

Finally, non-maxima suppression is performed and the edges are thresholded using hysteresis thresholding. We determine the hysteresis thresholds adaptively through examination of the histogram of gradient magnitudes.

Figure 2 shows an example of edge detection with and without stereo-guided scale selection. The original image has 750×500 pixels and can be found in Figure 1. In this example, the edges were detected at three scales ($\sigma = 1.0, 2.0, 4.0$) without the help of scale selection. Also given is the result with scale selection, where the filter response at each pixel was interpolated from the same three scales.

It can be seen that when a small scale ($\sigma = 1.0$) is used, many of the edges due to phenomena close to the camera are rough and a number of extraneous edges are detected due to the small scale, even though there is little image texture. However, when the scale is increased, we lose the details at the further phenomena (see, for example, the trees in the background and the end of the railing). On the other hand, when the scale is selected adaptively using the stereo range map, we have good performance at both close and far edge phenomena. In this case, the edges that are detected

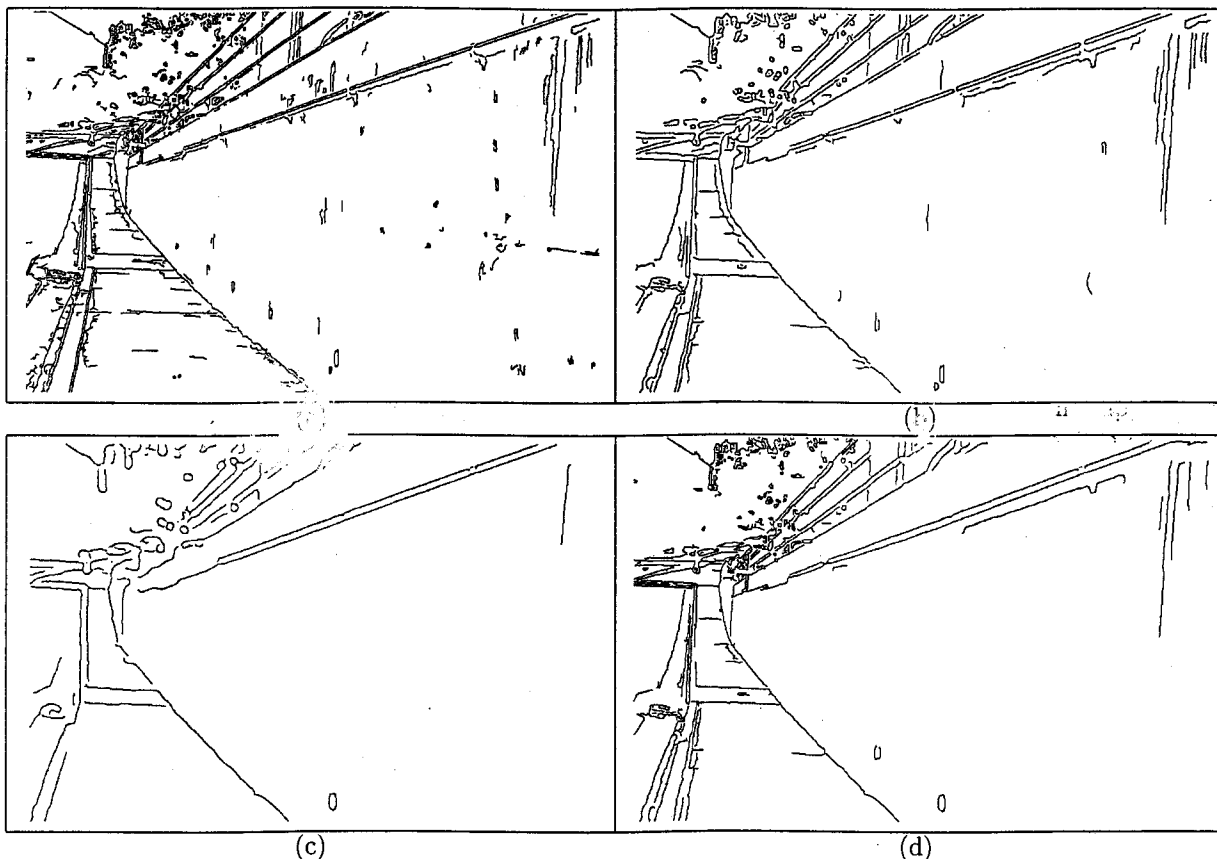


Figure 2: Edge detection results for the image in Figure 1. (a) Edges detected with $\sigma = 1.0$. (b) Edges detected with $\sigma = 2.0$. (c) Edges detected with $\sigma = 4.0$. (d) Edges detected with stereo-guided scale selection.

are at the same scale in world, rather than the same scale in the image.

5 Adaptive edge salience evaluation

In addition to its use in performing edge detection, range data is helpful in determining edge salience. Shorter edges that are detected at a larger distance are more likely to correspond to salient world edges than edges at close range that appear to be long due to perspective effects. We have primarily examined the summed gradient magnitude over the length of the edge and the local straightness of the edge as salience criteria, although many other salience measures could be used [14].

Consider, for example, a saliency measure where the gradient magnitude is summed along the length of the edge. The range data can be used to weight the gradient magnitude by the true edge length rather

than the image edge length.¹ Alternatively, we could sum the ranges to the pixels (normalized appropriately for the field-of-view and edge direction) to estimate the length of the edge in the world coordinates.

As a second example, we may consider the local straightness of an edge at each of its edge pixels by examining the difference in the gradient direction at neighboring edge pixels along the edge. However, we would not expect identical edge phenomena appearing at different ranges to yield the same differences in gradient direction between neighboring pixels. The edge closer to the camera will appear to be straighter locally, since the gradient differences will be smaller between neighboring pixels. To correct this situation, the differences in gradient direction can be weighted by the range to the edge.

We have implemented both of these techniques, and they have resulted in a substantial improvement in our

¹Note that, for non-frontal scenery, the orientation of the edge also affects the edge length. This effect can be accounted for if we estimate the three-dimensional orientation of the edge.

target application.

6 Relation to previous work

Our method for variable-scale smoothing can be interpreted as a technique to select, for each pixel in the image, a particular scale from the scale-space [15]:

$$S(x, y, \sigma) = I(x, y) * g(x, y, \sigma) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(u, v) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-u)^2 + (y-v)^2}{2\sigma^2}} du dv$$

We search for edges that occur near at the appropriate scale given by the stereo data and disregard the other scales.

Alternative methods for selecting a local scale from the scale-space have been given by several authors. Jeong and Kim [5] select the local scales through the minimization of an energy functional over the scale-space using a regularization approach. The functional includes terms that encourage a large scale in uniform intensity areas, a small scale where intensities change significantly, and a smoothly varying scale over the image. Morrone *et al.* [11] suggest that the local scale should be a monotonically decreasing function of the gradient magnitude. They argue that this results in good localization through the use of a small scale when the contrast is high and good sensitivity using a large scale with the contrast is low. Lindeberg [8] notes that edge detection procedures seek to find maxima in the gradient magnitude in the spatial variables and that this principal can also be applied to the scale variable. He thus seeks the edge position in the scale-space where gradient magnitude is maximized.

Unlike these methods, we select the local scale of examination based on an estimate of the true scale rather than trying to determine an appropriate scale through examination of the image. Our method is thus likely to yield better results when the real-world scale is the important one.

As an alternative to selecting a single scale, these techniques can be used to complement scale-space techniques [15]. In this case, the stereo range data would be used to transform the scale-space such that each scale plane was level with respect to the true scale rather than the image scale.

7 Results

Our target application for these techniques is to recognize surface-lying ordnance in military test ranges

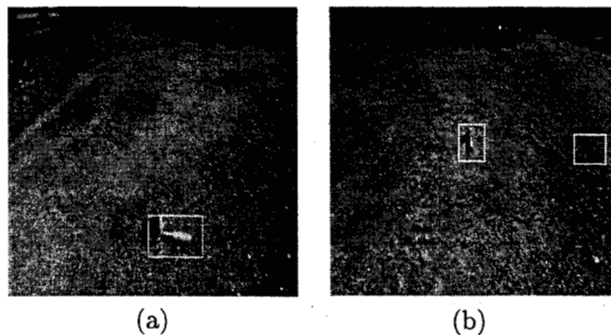


Figure 3: Ordnance recognition examples. (a) Correct detection at close range. (b) Correct detection at medium range and a false positive.

using a stereo system mounted on an unmanned ground vehicle for the purpose of autonomous remediation. One method to evaluate the edge detection techniques is by the performance of this application when using the stereo-guided smoothing and edge detection versus the performance when it is not used. We have tested the techniques using a set of 48 gray-scale images consisting of barren terrain with an inert piece of ordnance present at various distances and orientations (see Figure 3).

In this experiment, we tested three scales individually ($\sigma = 0.8, 1.6, 3.2$), the combination of all hypotheses found at the three discrete scales, and the result with stereo-guided scale selection using the same three scales to interpolate from. After edge detection was performed, an algorithm to detect the ordnance using geometric cues was used to find candidate positions [12].

Table 1 summarizes the results of this experiment. When the variable-scale smoothing and edge detection was performed, we achieved 40 correct recognitions out of the 48 cases. The eight failures occurred due to cases where the ordnance was a significant distance from the camera and at an orientation nearly aligned with the camera axis. In addition, 18 false positives were detected in the images. Figure 3 shows two examples, one of which contains a false positive.

For each individual scale that was examined, we had more cases where the ordnance was missed than with stereo-guided scale selection, and in two of them, we also found more false positives. While $\sigma = 4.0$ achieved 4 less false positives, the detection performance was significantly degraded, since 5 additional ordnance instances were missed. When all of the candidates from the three scales were combined (with duplicates removed), we achieved a slightly better detec-

Scale	False negatives	False positives
1.0	12	28
2.0	11	28
4.0	13	14
all	7	45
variable	8	18

Table 1: Results in ordnance recognition application.

tion performance with only 7 false negatives, but in this case the number of false positives rose sharply to 45.

Overall, the stereo selection techniques resulted in significantly superior performance to any of the individual scales or the combination of the scales.

8 Summary

We have described techniques that perform smoothing and edge detection adaptively using the results of stereoscopy to vary the scale at each pixel. This allows processing of the image to be performed with respect to the true scale of objects rather than the scale observed in the image. Stereoscopy has also been applied to evaluating the edge salience with respect to the true scale.

These techniques have been implemented as a variation of the Canny edge detector. We first convolve the image with Gaussian derivatives at a discrete set of scales. The correct response at each image pixel is estimated through parabolic interpolation of the known responses and normalization is performed so that the results are comparable across the image.

We have shown that these techniques yield desirable results on an image containing a wide range of scales. Furthermore, the application of this method to a data set for our target application of detecting unexploded ordnance in test ranges resulted in a considerable improvement in performance.

Acknowledgments

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, and was sponsored by the Air Force Research Laboratory at Tyndall Air Force Base, Panama City, Florida, through an agreement with the National Aeronautics and Space Administration.

References

- [1] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–697, November 1986.
- [2] W. T. Freeman and E. H. Adelson. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(9):891–906, September 1991.
- [3] D. B. Gennery. Least-squares camera calibration including lens distortion and automatic editing of calibration points. In A. Grün and T. S. Huang, editors, *Calibration and Orientation of Cameras in Computer Vision*. Springer-Verlag, in press.
- [4] E. R. Hancock and J. Kittler. Adaptive estimation of hysteresis threshold. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 196–201, 1991.
- [5] H. Jeong and C. I. Kim. Adaptive determination of filter scales for edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(5):579–585, May 1992.
- [6] J. J. Koenderink. The structure of images. *Biological Cybernetics*, 50:363–370, 1984.
- [7] T. Lindeberg. Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention. *International Journal of Computer Vision*, 11(3):283–318, 1993.
- [8] T. Lindeberg. Edge detection and ridge detection with automatic scale selection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 465–470, 1996.
- [9] L. Matthies. Stereo vision for planetary rovers: Stochastic modeling to near real-time implementation. *International Journal of Computer Vision*, 8(1):71–91, July 1992.
- [10] L. Matthies, A. Kelly, T. Litwin, and G. Tharp. Obstacle detection for unmanned ground vehicles: A progress report. In *Proceedings of the International Symposium on Robotics Research*, pages 475–486, 1996.
- [11] M. C. Morrone, A. Navangione, and D. Burr. An adaptive approach to scale selection for line and edge detection. *Pattern Recognition Letters*, 16:667–677, 1995.
- [12] C. F. Olson and L. H. Matthies. Visual ordnance recognition for clearing test ranges. In *Detection and Remediation Technologies for Mines and Mine-Like Targets III, Proc. SPIE*, 1998.
- [13] P. Perona. Deformable kernels for early vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(5):488–499, May 1995.
- [14] P. L. Rosin. Edges: Saliency measures and automatic thresholding. *Machine Vision and Applications*, 9:139–159, 1997.
- [15] A. P. Witkin. Scale-space filtering. In *Proceedings of the International Joint Conference on Artificial Intelligence*, volume 2, pages 1019–1022, 1983.
- [16] Y. Yakimovsky and R. Cunningham. A system for extracting three-dimensional measurements from a stereo pair of TV cameras. *Computer Vision, Graphics, and Image Processing*, 7:195–210, 1978.